Toxicology research for precautionary decision-making and the role of Human & Experimental Toxicology

Toxicology research for precautionary decision-making

and the role of Human & Experimental Toxicology

Philippe Grandjean, MD, PhD[1,2]

[1]Department of Public Health, University of Southern Denmark, J.B.Winslowsvej 17A, DK-5000 Odense C, Denmark

[2]Department of Environmental Health, Harvard School of Public Health, Landmark Center 3-110 East, 401 Park Drive, Boston, MA 02215

Corresponding author: P. Grandjean, Department of Public Health, University of Southern Denmark, J.B.Winslowsvej 17A**,** DK-5000 Odense, Denmark.

Telephone: +45 6550 3769

Email: pgrand@sdu.dk

Abstract

A key aim of toxicology is the prevention of adverse effects due to toxic hazards. Therefore, the dissemination of toxicology research findings must confront two important challenges: one being the lack of information on the vast majority of potentially toxic industrial chemicals, and the other being the strict criteria for scientific proof usually required for decision-making in regard to prevention. The present study ascertains the coverage of environmental chemicals in four volumes of *Human & Experimental Toxicology* and the presentation and interpretation of research findings in published articles. Links in SciFinder showed that the 530 articles published in four selected volumes between 1984 and 2014 primarily dealt with metals (126 links) and other toxicants that have received substantial attention in the past. Thirteen compounds identified by U.S. authorities in 2006 as high-priority substances, for which toxicology documentation is badly needed, were not covered in the journal issues at all. When reviewing published articles, reliance on *p* values was standard, and non-significant findings were often called "negative". This tradition may contribute to the perceived need to extend existing research on toxic hazards that have already been well characterised. Several sources of bias towards the null hypothesis can affect toxicology research, but are generally not considered, thus adding to the current inclination to avoid false positive findings. In this regard, toxicology is particularly prone to bias because of the known paucity of false positives and, in particular, the existence of a vast number of toxic hazards which by default are considered innocuous due to lack of documentation. The Precautionary Principle could inspire decision-making on the basis incomplete documentation and should stimulate a change in toxicology traditions and in toxicology research publication.

Keywords

Precaution, public health, research, scientific inference, toxicology

Introduction

Toxicology is the study of the adverse effects of chemical, physical or biological agents on living organisms and the ecosystem, including the prevention and amelioration of such adverse effects.[1] An important journal in the field, *Human & Experimental Toxicology* focuses on experimental and clinical studies of functional, biochemical and structural disorders and their causes, antidotes and other therapies. Toxicology research is in particular demand, as a substantial part of the total burden of disease in industrialized countries has been attributed to environmental factors, including chemicals,[2] and because documentation is lacking on the health effects of most industrial chemicals.[3] That said, one must ask if toxicology research and the dissemination of results through toxicology journals provide ample support for decision-making on preventing toxic hazards.

Risk assessment has so far relied on the so-called "untested-chemical assumption", i.e. that the lack of documentation means that regulatory action is not required.[3] This tradition has resulted in exposure limits for a small proportion of substances, and some limits were much too high to adequately protect against adverse health effects. For example, current limits for perfluorinated compounds in drinking water do not protect against their immunotoxic effects and may be more than 100-fold too high.[4] Thus, when scientific evidence is incomplete, exposure standards are more lenient, and prevention appears less important.

At least two major aspects deserve attention: the focus of the research and the interpretation of the results obtained.[5] Given the substantial number of untested chemical hazards, toxicology research would be expected to contribute documentation on the lesser-known hazards that have raised flags as potential toxicants. However, a recent bibliometric study of toxicology and public health journals showed that articles published during the first decade of

this millennium primarily addressed chemicals that had already been well studied, and that the top-10 substances were all metals.[6] Articles from *Human & Experimental Toxicology* up to 2009 were included in this bibliometric analysis, and the present study therefore aims to update the information through 2014.

While relying on toxicology research for guidance in regard to prevention needs, excessive demands for proof have prevented necessary regulation of toxic hazards.[3] The delay in decision-making may result in a continued, perhaps increasing, exposure to the suspected hazard. The Precautionary Principle (PP) has been proposed for situations of potentially serious or irreversible threats to health or to the environment, where the need to act to reduce potential hazards – before there is strong proof of harm – should take into account likely present and future costs and benefits of action and inaction.[7] Thus, preliminary, but reliable evidence may be sufficient to justify an intervention to avoid a health hazard that could otherwise lead to serious repercussions.

Demands for solid documentation before regulatory action will inspire continued elaboration of hazards that may already be well documented, since firm decisions are being deferred until current, perhaps minor, uncertainties can be resolved. Whereas, when decisions are PP-based, less extensive evidence would be required, as timely prevention of plausible harm would allow acceptance of some uncertainties as being inevitable or impossible to remove within at reasonable time frame. In the latter case, research attention can then be steered towards more poorly documented hazards that may pose new challenges and inspiration to toxicology research. Thus, a less extensive requirement regarding scientific evidence can have significant implications as to the ways that research is planned, performed, analysed, interpreted, and reported.[8] The present study analyses articles from *Human & Experimental Toxicology* to

ascertain how the journal and the toxicology research published in this journal fit in with current needs for toxicology documentation.

Methods

Subject matter

To assess the coverage of chemical substances in articles published in *Human & Experimental Toxicology* (and its predecessor, *Human Toxicology*), a bibliometric analysis was carried out using an internet-based database, SciFinder, that refers to individual substances by their Chemical Abstract Service (CAS) registry numbers.[6] Articles published in 2014 through to the October issue, as well as previous (complete) volumes published in 1984, 1994, and 2004 were scrutinized. The CAS numbers were translated into common names, and the frequencies were tabulated.

A separate search using SciFinder was made to ascertain to what extent thirteen high-priority environmental chemicals have been covered in *Human & Experimental Toxicology*. The thirteen high-production substances were identified by the U.S. Environmental Protection Agency in 2006 as lacking both a robust hazard data set and exposure information.[9] Despite the need for documentation of these compounds, they were only represented by a total of 352 links to scientific articles among the 180,000 articles published in the 72 journals during 2000-2009.[6] Additionally, very few articles on these compounds appeared during subsequent years.[5] The present study explored the coverage in *Human & Experimental Toxicology* volumes from 2004 and 2014; those substances that had been most frequently covered in toxicology journals were sought in the present journal in all volumes between 2006 and 2014.

Scientific inference

Toxicology research relies on statistical testing to assess whether or not an observed effect could be considered a possible result of natural variability. "Statistical significance" is achieved when the probability *p* is less than 5% for results that deviate at least as much as those observed from the expectation under the null hypothesis. However, when relying on the 5% limit, many potentially causal associations may inadvertently be ruled out, e.g. because the study was too small to reach statistical significance.[10] In addition, erroneous acceptance of the null hypothesis may be due to commonly occurring exposure misclassification, especially in human studies, but this bias is frequently ignored.[11] Similarly, several other weaknesses present in toxicology research can easily bias the findings towards false negative conclusions (Table 1).[7, 12] Nonetheless, the failure to reject the null hypothesis has often been interpreted as an indication of safety, although the lack of a statistically significant association alone in no way purports that there is an absence of an effect.[3] Accordingly, the scientific inference drawn from toxicology research needs to judge the *p* values in light of the study characteristics.

Prudent interpretation of toxicology research requires that 1) the sources of uncertainty and potential underestimation of risk are identified and taken into account; 2) the results are presented with statistical confidence limits; 3) the results are interpreted in light of the statistical power of the study; and 4) the upper confidence limit is considered as part of a worst-case scenario when interpreting the results. These questions were used to evaluate the selected *Human & Experimental Toxicology* articles. To limit the extent of this evaluation, articles published in the years 1984, 1994, 2004, and 2014 were selected, starting with the first issue of each year. Each article was independently evaluated by two assessors, and any discrepancies were resolved by consensus.

Results

The four selected volumes of the journal included 530 articles, from which SciFinder retrieved a total of 541 different CAS numbers, of which a total of 122 were covered at least four times. However, this number includes substances of limited toxicity, such as water, normal components of the body, such as glutathione (45 articles), ascorbic acid (13), and a variety of enzymes. The industrial chemicals, excluding drugs, most commonly covered in *Human & Experimental Toxicology* are listed in Table 2. Metals are clearly in frequent focus (with a total of 126 article links), as are ethanol and certain pesticides, such as lindane and paraquat, though not in the more recent volumes. Of the 27 substances listed in Table 2, 15 are elements, mostly metals, but the list also includes five pesticides and three solvents (including ethanol). Among pharmaceuticals not included in the table, cisplatin (16), acetaminophen (8), propoxyphene (6), and acetylsalicylic acid (6) were the most popular.

When searching for the U.S.EPA's list of 13 substances urgently requiring toxicological and exposure information,[9] the SciFinder data indicated that none of them were reported at all in the four journal volumes examined. Additional searches were made for four substances that had been covered in at least ten articles in toxicology journals during 2000-2009,[6] i.e., 1,3-dichlorobenzene, bromochloromethane, triclocarban, and hexabromocyclododecane. Again, none of them was found in the *Human & Experimental Toxicology* volumes published between 2006 and 2014.

In exploring how the results were presented and interpreted, the initial strategy turned out to be too ambitious, and a qualitative presentation of results seemed most appropriate. A thorough review was restricted to the first issue of each volume, but subsequent issues from the same year were also screened. In regard to the first question asked, almost all articles highlighted

uncertainties to some extent. However, almost all presented solely *p* values and relied on those when interpreting the results, usually without clear emphasis on limitations due to statistical power and other limitations. Standard deviations were often provided, especially in more recent articles, but confidence limits were generally not provided. In the discussion section, the articles reviewed frequently referred to results being "negative" if a significant difference ($p < 0.05$) was not observed. The findings were similar for environmental chemicals and drugs, and there appeared to be no obvious time trend.

Discussion

Human & Experimental Toxicology is a highly respected toxicology journal, and published articles are frequently cited, thus reflecting the journal's high standing. The findings of the present study likely addresses publication practises that are typical of modern toxicology research. In agreement with a previous study of 72 toxicology and public health journals,[6] the industrial chemicals most frequently covered in articles published in this journal were mainly ones commonly studied in the past and that continue to be featured in the toxicology literature.

While more detailed exploration and verification of some of these well-studied substances may be possible to justify, such choices must be considered in the light of the concomitant lack of coverage of substances for which an urgent need for research has been announced. The present study therefore supports the existence of substantial inertia that favours research and publication on a limited number of substances rather than poorly-known chemicals for which documentation is in demand.

An important reason for such inertia relates to the science paradigm that requires replication and verification to justify solid conclusions. In support of this tradition, many

preliminary findings that were highly publicized later proved to be wrong.[13] As a result, both funding agencies[14] and journals[15] have announced their intention of increasing reproducibility of research. Although crucial to support the credibility of research, emphasis on replication can also have untoward effects, as certain environmental chemicals already trigger over 1,000 publications per year.[6] In this case, replication seems to be pursued to an extreme and perhaps counterproductive extent by merely extending the current focus on well-known hazards in toxicology. This phenomenon is well known in science and has been dubbed the "Matthew" effect (referring to the New Testament: "For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath").[16] However, from an innovation point of view, and from the point of view of the PP, the opposite strategy, e.g. targeting newly recognized potential hazards or researching lesser known toxicants, would appear much more attractive.

While research institutions, funding agencies, and the research community share a substantial part of the responsibility for the inertia, toxicology journals may too. Manuscripts on well-known substances are easier to find peer reviewers for, and they may be more likely to be cited. Perhaps some environmental chemicals are held in higher esteem than others, thereby adding to their continued prominence, or publication persistence, no matter what the societal needs may be. This means that there may be an element of circular reasoning involved, where a substance is a popular research item simply because it has been widely studied in the past – a self-prophetic bias that maintains a continued prominence of a specific scientific community and its publications.[16] Science journals could likely play a role in inspiring change. Thus, otherwise excellent research may not represent a top priority for publication if it merely extends

unrestrained replication of well-known phenomena, and editors could instead assign higher priority to studies that explore new territory.

A U.S. National Research Council committee recently called attention to the erroneous inference that chemicals are regarded inert or safe, unless proven otherwise.[3] Thus, inconclusive studies have sometimes been labelled as "negative" or were thought to represent "no risk" rather than "uncertain information".[17] This tendency is clear also in the present journal, and it likely amplifies the tendency towards continuing research on particular substances, as uncertainties will always prevail on some aspect of their toxicological properties, thus appearing to justify continued research on this subject.

This tradition appears even more misleading when the 5% limit is strictly applied, so that a greater emphasis is placed on results that have a *p* value of, say, 4.9% than on results with one of 5.1%,[18] although there is of course no meaningful difference between outcomes with such similar p values. Adding further misperception, the null hypothesis is usually that an exposure has no effect. Thus, in toxicology research, the *p* value is frequently used to test a null hypothesis that may be unrealistic or obviously wrong, i.e. that lead exposure is not neurotoxic. A solution may be to test the results against a plausible alternative hypothesis, but such hypothesis may not be sufficiently specific to allow testing by Bayesian statistics.[19]

Some scientists and some scientific journals oppose the reliance on *p* values.[20] Instead of calculating whether the point estimate is "significantly" different from no effect, the confidence interval should be preferred.[21] It represents the range of values within which 95% of effect estimates would likely fall if a large number of similar studies were conducted. In other words, given the point estimate and the calculated variability, the study would not be in disagreement with any hypothesis that postulated an effect within the confidence interval. If the interval

includes no effect or no difference, the deviation observed has not reached statistical significance. Still, the results can also be said to be in accordance with many other hypotheses, some perhaps suggesting a serious effect or a large difference. Such alternative interpretations are usually ignored, although the upper confidence limit may represent a plausible worst case scenario that should deserve attention. In agreement with this notion, the extreme confidence limit for the benchmark dose is used for risk assessment purposes.[22]

The present review of articles from *Human & Experimental Toxicology* suggests that *p* values are still widely favoured and that the range of the confidence interval and its implications are rarely, if ever, considered. Further, the many factors listed in Table 1 that may bias the study findings towards the null are often ignored, thereby adding to the tendency towards underestimating toxic effects. Although other journals were not included in this study, *Human & Experimental Toxicology* most likely reflects tendencies that are typical for scientific journals in the field.

Still, the risk of false positives should also be considered. Thus, within certain fields of biomedical research, a large proportion of published conclusions are claimed to be false.[13] Although that may be true for certain fields of research particularly affected by publication bias, the field of toxicology is different, especially in regard to the large number of industrial chemicals. In toxicology, the ratio of true to no relationships among the relationships probed in published studies is likely much higher than in most other fields. Thus, only a small percentage of industrial chemicals in use in the late 1970es were considered hazardous, while that was true for about 70% of new chemicals tested.[23] The "untested chemicals assumption" therefore causes a very large proportion of false *negative* conclusions. In contrast, when scrutinising alleged false *positive* findings in environmental health and toxicology, very few such cases have been found.[24]

Thus, although publication bias could potentially cause a tendency of favouring false positive results in toxicology as in other fields, the impact is negligible in comparison with the huge number of environmental chemicals, for which virtually no toxicology information exists.

Given the uneven research coverage of toxic hazards, interpretation of the evidence needs to address the following question:[5] What could possibly be known, given the type of evidence available? Noisy studies, e.g. with imprecise estimates of the causative exposure and insensitive and nonspecific outcome measures, are likely to detect only the most serious risks. They should be cautiously interpreted in light of their (limited) weight of evidence. The fact that the null hypothesis could not be rejected with confidence may be irrelevant in such cases, and any support for the absence of toxicity would be very small. These considerations are of particular importance also in a wider perspective, as they are usually not dealt with in recommendations for systematic review.[25-27]

The present study suggests that toxicology, as revealed by publications in *Human & Experimental Toxicology*, to some extent appears to fail the purpose of providing documentation needed by society in regard to preventing toxic hazards. Toxicology should not be credulous and overly generous accepting potentially causal associations so that it prevents the distinction between useful and misguided ideas. Still, inherent biases toward the null need to be recognized, and the logical error of assuming safety when a hazard could not be convincingly demonstrated in a "negative" study needs to be thwarted. The PP would seem to represent a proper antidote to this adverse tendency in toxicology.[17]

Toxicology documentation must be acknowledged as provisional and dynamic. As the true extent of toxicity caused by industrial chemicals is poorly known, current knowledge most likely underestimates the adverse effects to a very considerable degree.[7] Although this situation

might call for prudent precaution when interpreting toxicity data, the present study illustrates that a serious bias exists in the opposite direction, both in regard to the coverage of chemical hazards and in regard to the interpretation of research results. Toxicology information and risk assessment can never be complete, but can become less biased. Hence, there is room for improvement. While journal editors have little power to influence research planning, they can guide and inspire the reporting of results, thereby further improving the status and usefulness of toxicology research. Thus, this study suggests an additional and important role for *Human & Experimental Toxicology* to fill in the years to come.

Acknowledgment

Katherine T. Herz and Constance G. Poulsen helped to evaluate the *Human & Experimental Toxicology* articles, and Esben Budtz-Jørgensen provided statistical advice.

Conflict of interest

The author declares no conflicts of interest.

Funding

This study received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Table 1. Examples of features that may bias toxicology toward missing a true association (false negative).[5, 7]

Toxicological studies in general
Low statistical power
Use of 5 % probability level
Pressures against false alarm
Epidemiological studies of toxicants:
Inappropriate control group
Exposure misclassification
Inadequate follow-up of exposed subjects (cases lost to follow-up, follow-up too short)
Experimental toxicology:
Exposure to single substances, one at the time
Limited number of dose levels
Exposure duration less than lifetime
Standard, non-specific effect measures
Inbred strains to limit genetic variability

Table 2. Industrial chemicals covered in at least five articles in the four volumes (total number of articles published in parenthesis).

Name	CAS#	1984 (203)	1994 (136)	2004 (83)	2014 (108)	Total (530)
Ethanol	64-17-5	6	6	6	2	20
Mercury	7439-97-6	7	7	2		16
Copper	7440-50-8	5	5	2	1	13
Lead	7439-92-1	4	4	3	2	13
Nickel	7440-02-0	6	6			12
Zinc	7440-66-6	5	5		1	11
Lindane	58-89-9	5	5			10
Paraquat	4685-14-7	5	5			10
Acrolein	107-02-8	3	3	2		8
Cadmium	7440-43-9	4	4			8
Mustard gas	505-60-2	4	4			8
Nitric oxide	10102-43	1	1	3	3	8
Iron	7439-89-6			5	2	7
Arsenic	7440-38-2	3	3	1		7
Manganese	7439-96-5	3	3		1	7
Aluminum	7429-90-5	2	2	2		6
Chromium	7440-47-3	3	3			6
Ethylene glycol	107-21-1	3	3			6
Fluazifop-butyl	69806-50	3	3			6
Nitrogen dioxide	10102-44	3	3			6
Selenium	7782-49-2			4	1	5
p,p′-DDT	50-29-3	1	1	2	1	5
Carbon tetrachloride	56-23-5	2	2	1		5
Cotinine	486-56-6	2	2		1	5
Cypermethrin	52315-07	2	2		1	5
Lithium	7439-93-2	2	2	1		5
Magnesium	7439-95-4	2	2		1	5
Ozone	10028-15	2	2		1	5
Platinum	7440-06-4	2	2		1	5

References

1. Society of Toxicology. Communiqué: How do you define toxicology? 2005.
2. Pruss-Ustun A, Vickers C, Haefliger P, et al. **Knowns and unknowns on burden of disease due to chemicals: a systematic review**. *Environ Health* 2011, **10:** 9.
3. National Research Council. *Science and decisions: advancing risk assessment*. Washington, D.C.: National Academy Press, 2009.
4. Grandjean P and Budtz-Jorgensen E. **Immunotoxicity of perfluorinated alkylates: calculation of benchmark doses based on serum concentrations in children**. *Environ Health* 2013, **12:** 35.
5. Grandjean P. Science for precautionary decision-making. In: Gee D, Grandjean, P., Hansen, S.F., van den Hove, S., MacGarvin, M., Martin, J., Nielsen, G., Quist, D., Stanners, D., (ed.). *Late Lessons from Early Warnings*. Copenhagen: European Environment Agency, 2013, p. 517-535.
6. Grandjean P, Eriksen ML, Ellegaard O, et al. **The Matthew effect in environmental science publication: a bibliometric analysis of chemical substances in journal articles**. *Environ Health* 2011, **10:** 96.
7. Grandjean P. **Implications of the precautionary principle for primary prevention and research**. *Ann Rev Publ Health* 2004, **25:** 199-223.
8. Grandjean P. **Late insights into early origins of disease**. *Basic Clin Pharmacol Toxicol* 2008, **102:** 94-99.
9. Environmental Protection Agency (EPA). Risk-Based Prioritization (RBP) Decisions Summary. 2009.
10. Goodman S. **A dirty dozen: twelve p-value misconceptions**. *Semin Hematol*. 2008, **45:** 135-140.
11. Grandjean P and Budtz-Jorgensen E. **An ignored risk factor in toxicology: The total imprecision of exposure assessment**. *Pure Appl Chem*. 2010, **82:** 383-391.
12. Gee D. **Establishing evidence for early action: the prevention of reproductive and developmental harm**. *Basic Clin Pharmacol Toxicol* 2008, **102:** 257-266.
13. Ioannidis JP. **Why most published research findings are false**. *PLoS Med* 2005, **2:** e124.
14. Collins FS and Tabak LA. **Policy: NIH plans to enhance reproducibility**. *Nature* 2014, **505:** 612-613.
15. **Journals unite for reproducibility**. *Nature* 2014, **515:** 7.
16. Merton RK. **The Matthew effect in science. The reward and communication systems of science are considered**. *Science* 1968, **159:** 56-63.
17. Grandjean P. **Seven deadly sins of environmental epidemiology and the virtues of precaution**. *Epidemiology* 2008, **19:** 158-162.
18. Holman CD, Arnold-Reed DE, de Klerk N, et al. **A psychometric experiment in causal inference to estimate evidential weights used by epidemiologists**. *Epidemiology* 2001, **12:** 246-255.
19. Johnson VE. **Revised standards for statistical evidence**. *Proc Nat Acad Sci U.S.A.* 2013, **110:** 19313-19317.
20. Lang JM, Rothman KJ and Cann CI. **That confounded P-value**. *Epidemiology* 1998, **9:** 7-8.
21. Thompson WD. **Statistical criteria in the interpretation of epidemiologic data**. *Am J Public Health* 1987, **77:** 191-194.

22. EFSA Scientific Committee. **Guidance of the Scientific Committee on Use of the benchmark dose approach in risk assessment**. *The EFSA Journal*. 2009, **1150:** 1-72.
23. Denison RA. Not that innocent: A comparative analysis of Canadian, European Union and United States policies on industrial chemials, 2007.
24. Hansen SF, Krayer von Krauss MP and Tickner JA. **Categorizing mistaken false positives in regulation of human and environmental health**. *Risk Anal* 2007, **27:** 255-269.
25. Guyatt G, Oxman AD, Akl EA, et al. **GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables**. *J Clin Epidemiol* 2011, **64:** 383-394.
26. Rooney AA, Boyles AL, Wolfe MS, et al. **Systematic review and evidence integration for literature-based environmental health science assessments**. *Environ Health Perspect* 2014, **122:** 711-718.
27. Woodruff TJ and Sutton P. **The Navigation Guide systematic review methodology: a rigorous and transparent method for translating environmental health science into better health outcomes**. *Environ Health Perspect* 2014, **122:** 1007-1014.